C.S. A1: Analysis of Patient Data from Secondary Sources

C.S. A1.1: Secondary Data with Links to SSNs

Overview

This study uses analysis of patient data from existing VA databases (originally established for patient care and administrative purposes—not research) to compare statistical models of risk adjustment and mortality prediction. There is no direct patient contact, and scrambled patient identifiers are used to link data from various sources. (See additional explanation below under "Data Collection and Confidentiality.")

Subjects and Sample Size

Data are collected on 5,000 VA patients with ICU admissions. Subjects are to be identified from VA databases using diagnostic criteria.

Data Collection and Confidentiality

Patient data to be collected include demographic information, date of birth, zip code, gender, ethnicity, ICU admissions, diagnoses, lab results, inpatient treatment information, mortality data, and other outcomes. These data are collected via database search (e.g., Austin data, Pharmacy Benefits Management data, DSS data) and will be used to test and compare risk adjustment methods.

The patient cohort will be obtained from existing VA databases using diagnostic criteria. Some of the databases contain real SSNs, others contain scrambled SSNs. After the study data, including SSNs and scrambled SSNs, are pulled, all real SSNs will be converted to scrambled SSNs, using a file linking scrambled SSNs with real SSNs obtained from a separate Austin database. Thus, all study files with patient data will include only scrambled SSNs. The file linking scrambled SSNs with real SSNs will be maintained by the research team as a separate file, in a password-protected drive that is separate from the drive containing the study data.

C.S. A1.1

[From OHRP Web site: www.hhs.gov/ohrp/humansubjects/guidance/decisioncharts.htm]

What type of review should this application receive and why?

Chart 2: Is the research involving human subjects eligible for exemption under 45 CFR 46.101(b)?

Will the ***only*** involvement of human subjects be in one or more of the following categories?

- Research conducted in established or commonly accepted educational settings, involving normal education practices? OR
- Research involving the use of educational tests, survey procedures, interview procedure, or observation of public behavior? OR
- Research involving collection or study of ***existing*** data, documents, records, or pathological or diagnostic specimens? OR
- Research studying, evaluating, or examining public benefit or service programs? OR
- Research involving taste and food quality evaluation or consumer acceptance studies?

↓

YES: Research involving collection or study of ***existing*** data, documents, records, or pathological or diagnostic specimens.

↓

Exemption 45 CFR 46.101(b)(4) may apply.
Go to Chart 5.

↓

Chart 5

↓

Are these sources publicly available?

↓

NO

↓

Will information be ***recorded by the investigator*** in such a manner that the subjects ***cannot be identified***, directly or through identifiers linked[1] to the subjects?

↓

NO

↓

Research is not exempt.
Go to Chart 8.

↓

Chart 8

↓

Has the research been ***previously reviewed*** an approved by the IRB?

↓

NO

↓

Does the research present ***no more than minimal risk***[2] to human subjects?
AND
Does the research involve ***only procedures included in categories 1 through 7***[3]
on the list of categories that may be reviewed through an expedited review procedure? [45 CFR 10(b)(1)]

↓

YES

↓

Is the research classified?

↓

NO

↓

Research is eligible for IRB review through expedited procedures. [45 CFR 46.110(d)]

Notes for C.S. A1

[1]**Definition:** There are identifiers in the study data set that can be linked to the subject.

[2]**Definition:** "Minimal risk means that the ***probability*** and ***magnitude*** of harm or discomfort anticipated in the research are not greater in and of themselves than those ***ordinarily encountered in daily life*** or during the performance of routine physical or psychological examinations or tests" (CFR 46.102(1)).

Discussion: Panel members considered this study to be minimal risk. The ***probability and magnitude*** of loss of confidentiality, given the safeguards described, are no greater than ***that which is encountered in daily life***—e.g., the probability of loss of confidentiality of other, non-research-related health data collected and maintained within VA medical centers and clinics, or by non-VA healthcare providers. The small probability of loss of confidentiality is based on the assumption that the safeguards for maintaining data confidentiality by the investigators are adequate—or, at a minimum, that the procedures are as good as those used elsewhere in the health care facility for ensuring the confidentiality of health-related data. Therefore, sufficient information must be provided by investigators to the IRB committee for them to determine that the procedures for maintaining data confidentiality are acceptable. If there has been a history of problems with maintaining the confidentiality of research data at a particular institution, or if the investigators do not have much experience with the collection and use of data from secondary datasets, then the local IRB may choose full review as a means to more carefully review the procedures and ensure that they are adequate.

[3]**Definition:** The research involves procedures included in category 5: Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis). [Return to home page for full list of categories eligible for expedited review under 45 CFR 46.110(b)(1).]

C.S. A1.2: Secondary Data with Scrambled SSNs

Overview

This study uses analysis of patient data from existing VA databases (originally established for patient care and administrative purposes—not research) to compare statistical models of risk adjustment and mortality prediction. There is no direct patient contact, and scrambled patient identifiers are used to link data from various sources. (See additional explanation below under "Data Collection and Confidentiality.")

Subjects and Sample Size

Data are collected on 5,000 VA patients with ICU admissions. Subjects are to be identified from VA Austin databases using diagnostic criteria.

Data Collection and Confidentiality

Patient data to be collected include demographic information, date of birth, diagnoses, procedures, inpatient service use, mortality, and outpatient service use. These data are collected from national (Austin) VA databases, including the inpatient and outpatient databases.

The patient cohort will be obtained using diagnostic criteria. Data obtained from these databases will include scrambled SSNs, which are needed to link patients across time and across multiple databases. The scrambled SSN for a given VA patient is the same across all Austin databases; so, the scrambled identifier can be used to link data for a given patient across databases.

All study files with patient data will include only scrambled SSNs. The file linking scrambled SSNs with real SSNs is in a database maintained in Austin.

C.S. A1.2

[From OHRP Web site: www.hhs.gov/ohrp/humansubjects/guidance/decisioncharts.htm]

What type of review should this application receive and why?

Chart 2: Is the research involving human subjects eligible for exemption under 45 CFR 46.101(b)?

Will the ***only*** involvement of human subjects be in one or more of the following categories?

- Research conducted in established or commonly accepted educational settings, involving normal education practices? OR
- Research involving the use of educational tests, survey procedures, interview procedure, or observation of public behavior? OR
- Research involving collection or study of ***existing*** data, documents, records, or pathological or diagnostic specimens? OR
- Research studying, evaluating, or examining public benefit or service programs? OR
- Research involving taste and food quality evaluation or consumer acceptance studies?

↓

YES: Research involving collection or study of ***existing*** data, documents, records, or pathological or diagnostic specimens.

↓

Exemption 45 CFR 46.101(b)(4) may apply.
Go to Chart 5.

↓

Chart 5

↓

Are these sources publicly available?

↓

NO

↓

Will information be ***recorded by the investigator*** in such a manner that the subjects ***cannot be identified***, directly or through identifiers linked[1] to the subjects?

↓

NO

↓

Research is not exempt.
Go to Chart 8.

↓

Chart 8

↓

Has the research been ***previously reviewed*** an approved by the IRB?

↓

NO

↓

Does the research present ***no more than minimal risk***[2] to human subjects?
AND
Does the research involve ***only procedures included in categories 1 through 7***[3]
on the list of categories that may be reviewed through an expedited review procedure? [45 CFR 10(b)(1)]

↓

YES

↓

Is the research classified?

↓

NO

↓

Research is eligible for IRB review through expedited procedures. [45 CFR 46.110(d)]

Notes for C.S. A1.2

[1]**Definition:** There are identifiers in the study data set that can be linked to the subject. The linkage is maintained in another database in another computer in a site separate from where the research is taking place. The investigator has access to that database, so, the investigator knows the identity of the human subjects. In addition, the specific data that are being obtained from the dataset may themselves identify individuals, even if SSN or name is not obtained.

Discussion: While the majority of the panel members felt that there was the possibility that human subjects could be identified from the dataset, several panel members felt that because the investigator was not recording or maintaining the linkage, the answer to this question is "YES"—and, therefore, the study should be considered exempt from human subjects review. However, in order to be exempt in this case, the investigator would also need to provide information showing that none of the other data obtained from the dataset could be used to identify individuals.

[2]**Definition:** "Minimal risk means that the ***probability*** and ***magnitude*** of harm or discomfort anticipated in the research are not greater in and of themselves than those ***ordinarily encountered in daily life*** or during the performance of routine physical or psychological examinations or tests" (CFR 46.102(1)).

Discussion: Panel members considered this study to be minimal risk. The ***probability and magnitude*** of loss of confidentiality, given the safeguards described, are no greater than ***that which is encountered in daily life***—e.g., the probability of loss of confidentiality of other, non-research-related health data collected and maintained within VA medical centers and clinics, or by non-VA healthcare providers. The small probability of loss of confidentiality is based on the assumption that the safeguards for maintaining data confidentiality by the investigators are adequate—or, at a minimum, that the procedures are as good as those used elsewhere in the health care facility for ensuring the confidentiality of health-related data. Therefore, sufficient information must be provided by investigators to the IRB committee for them to determine that the procedures for maintaining data confidentiality are acceptable. If there has been a history of problems with maintaining the confidentiality of research data at a particular institution, or if the investigators do not have much experience with the collection and use of data from secondary datasets, then the local IRB may choose full review as a means to more carefully review the procedures and ensure that they are adequate.

[3]**Definition:** The research involves procedures included in category 5: Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis). [Return to home page for full list of categories eligible for expedited review under 45 CFR 46.110(b)(1).]

C.S. A1.3: Use of SSNs to Obtain Secondary Data

Overview

This study uses analysis of patient data from existing VA databases (originally established for patient care and administrative purposes—not research) and from electronic medical records to compare statistical models of risk adjustment and mortality prediction. There is no direct patient contact, and scrambled patient identifiers are used to link data from various sources. (See additional explanation below under "Data Collection and Confidentiality.")

Subjects and Sample Size

Data are collected on 5,000 VA patients with ICU admissions. Subjects are to be identified from VA Austin databases using diagnostic criteria.

Data Collection and Confidentiality

Austin databases are searched by diagnostic code to determine eligible patients. The SSNs of these eligible patients are obtained, and are used to identify the patients' electronic medical records, from which relevant clinical data for the study are obtained. These clinical data are entered into the study database, which does not contain any SSNs or linkages to patient identifiers. A file containing the list of SSNs of eligible patients—but no linkages and no other data—is temporarily maintained until all of the medical record data are obtained.

C.S. A1.3

[From OHRP Web site: www.hhs.gov/ohrp/humansubjects/guidance/decisioncharts.htm]

What type of review should this application receive and why?

Chart 2: Is the research involving human subjects eligible for exemption under 45 CFR 46.101(b)?

Will the ***only*** involvement of human subjects be in one or more of the following categories?

- Research conducted in established or commonly accepted educational settings, involving normal education practices? OR
- Research involving the use of educational tests, survey procedures, interview procedure, or observation of public behavior? OR
- Research involving collection or study of ***existing*** data, documents, records, or pathological or diagnostic specimens? OR
- Research studying, evaluating, or examining public benefit or service programs? OR
- Research involving taste and food quality evaluation or consumer acceptance studies?

↓

YES: Research involving collection or study of ***existing*** data, documents, records, or pathological or diagnostic specimens.

↓

Exemption 45 CFR 46.101(b)(4) may apply.
Go to Chart 5.

↓

Chart 5

↓

Are these sources publicly available?

↓

NO

↓

Will information be ***recorded by the investigator*** in such a manner that the subjects ***cannot be identified***, directly or through identifiers linked[1] to the subjects?

↓

NO

↓

Research is not exempt.
Go to Chart 8.

↓

Chart 8

↓

Has the research been ***previously reviewed*** an approved by the IRB?

↓

NO

↓

Does the research present ***no more than minimal risk***[2] to human subjects?
AND
Does the research involve ***only procedures included in categories 1 through 7***[3]
on the list of categories that may be reviewed through an expedited review procedure? [45 CFR 10(b)(1)]

↓

YES

↓

Is the research classified?

↓

NO

↓

Research is eligible for IRB review through expedited procedures. [45 CFR 46.110(d)]

Notes for C.S. A1.3

[1]**Discussion:** The majority of the panel felt that the use of SSNs means there are linkages to data (i.e., the study is not exempt). Several panel members felt that because the investigator is not linking the SSNs with study numbers or study data, that there are no linkages, and the study should be considered exempt from human subjects review. However, in order to be exempt in this case, the investigator would also need to provide information showing that none of the other data obtained from the dataset could be used to identify individuals.

[2]**Definition:** "Minimal risk means that the ***probability*** and ***magnitude*** of harm or discomfort anticipated in the research are not greater in and of themselves than those ***ordinarily encountered in daily life*** or during the performance of routine physical or psychological examinations or tests" (CFR 46.102(1)).

Discussion: Panel members considered this study to be minimal risk. The ***probability and magnitude*** of loss of confidentiality, given the safeguards described, are no greater than ***that which is encountered in daily life***—e.g., the probability of loss of confidentiality of other, non-research-related health data collected and maintained within VA medical centers and clinics, or by non-VA healthcare providers. The small probability of loss of confidentiality is based on the assumption that the safeguards for maintaining data confidentiality by the investigators are adequate—or, at a minimum, that the procedures are as good as those used elsewhere in the health care facility for ensuring the confidentiality of health-related data. Therefore, sufficient information must be provided by investigators to the IRB committee for them to determine that the procedures for maintaining data confidentiality are acceptable. If there has been a history of problems with maintaining the confidentiality of research data at a particular institution, or if the investigators do not have much experience with the collection and use of data from secondary datasets, then the local IRB may choose full review as a means to more carefully review the procedures and ensure that they are adequate.

[3]**Definition:** The research involves procedures included in category 5: Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis). [Return to home page for full list of categories eligible for expedited review under 45 CFR 46.110(b)(1).]